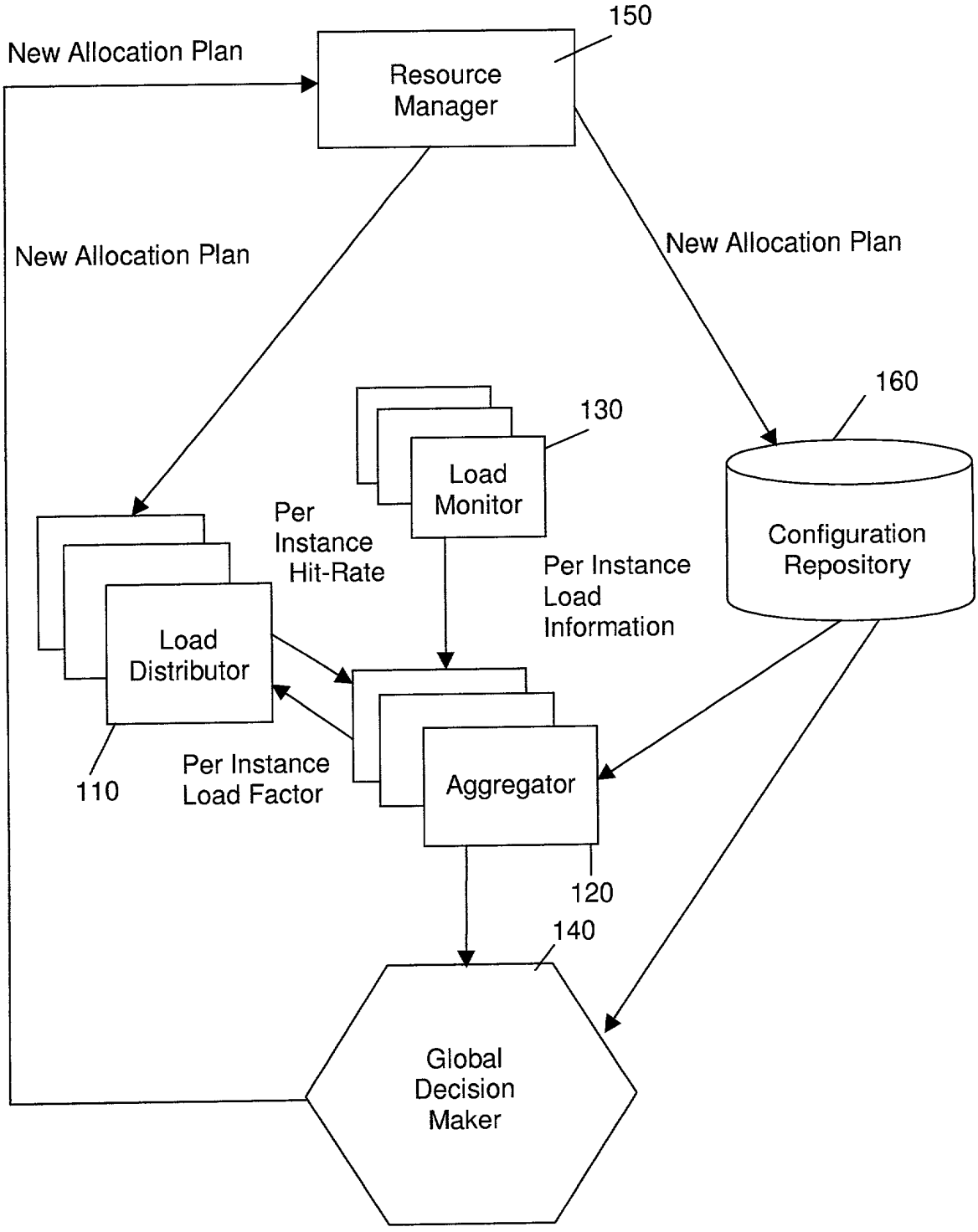
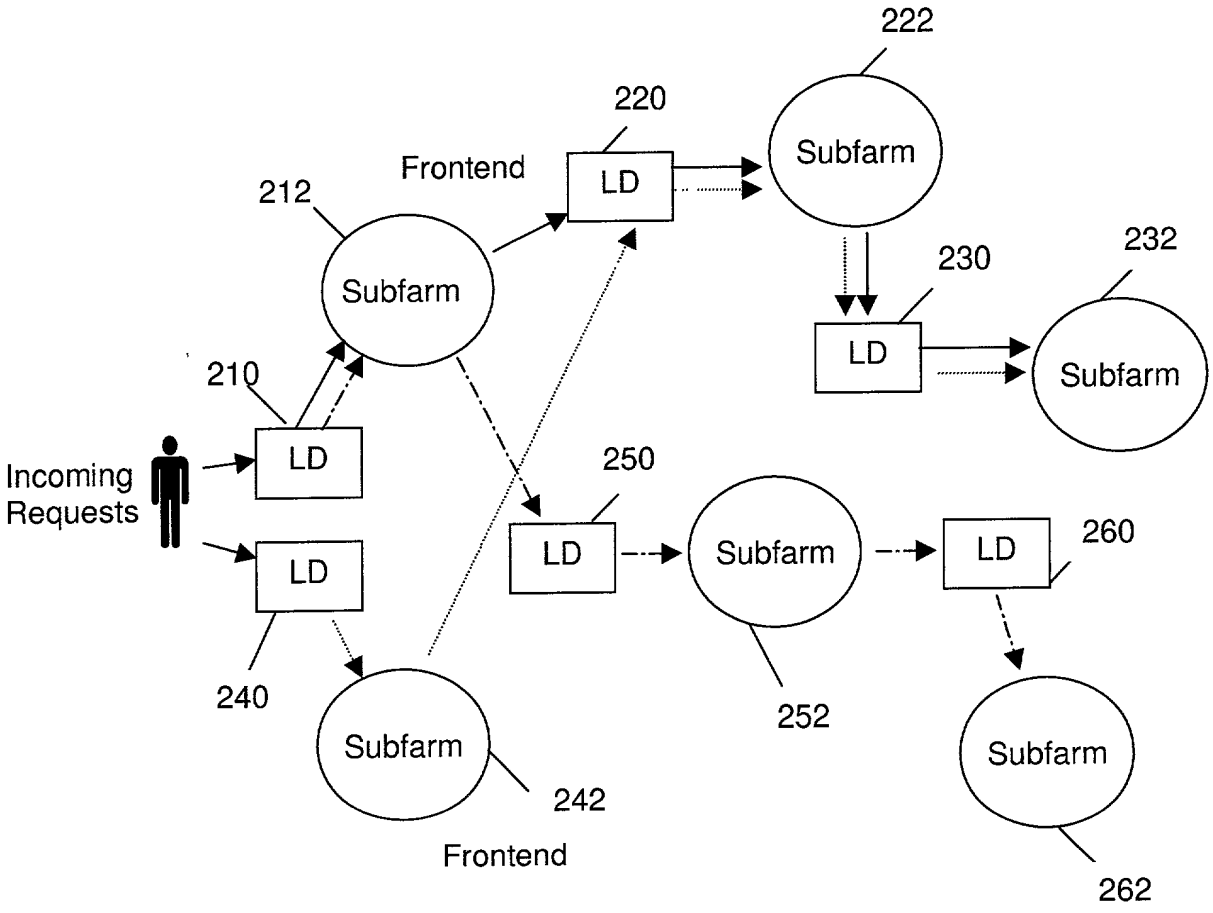

150
New Allocation Plan
Resource Manager
New Allocation Plan
New Allocation Plan
160
130
Load Monitor
Per Instance Hit-Rate
Configuration Repository
Per Instance Load Information
Load Distributor
110
Per Instance Load Factor
Aggregator
120
140
Global Decision Maker

FIG 1



Request for hosted application 1
Request for hosted application 2
Request for hosted application 3

LD: Load Distributor for sub-farm

FIG 2

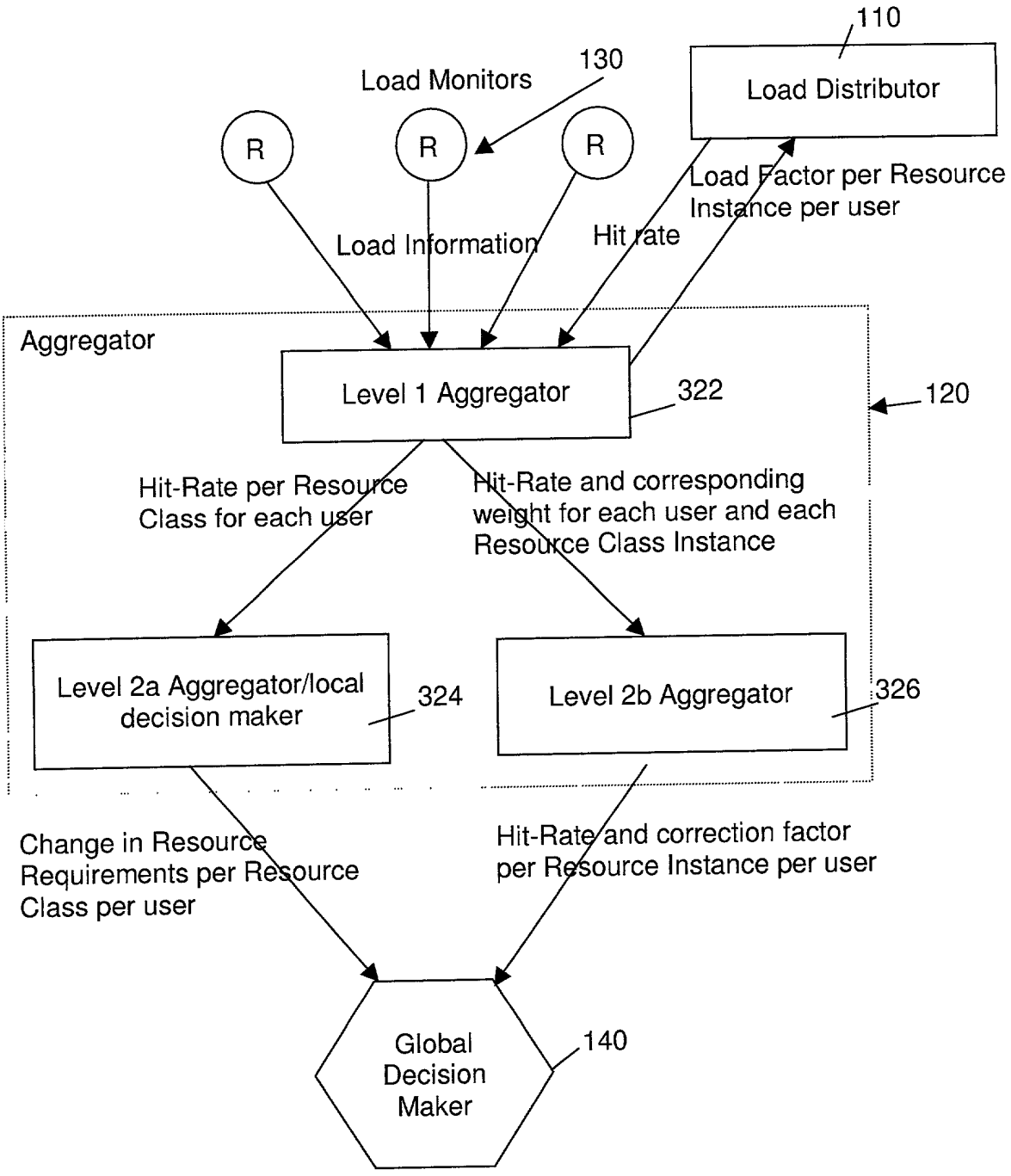
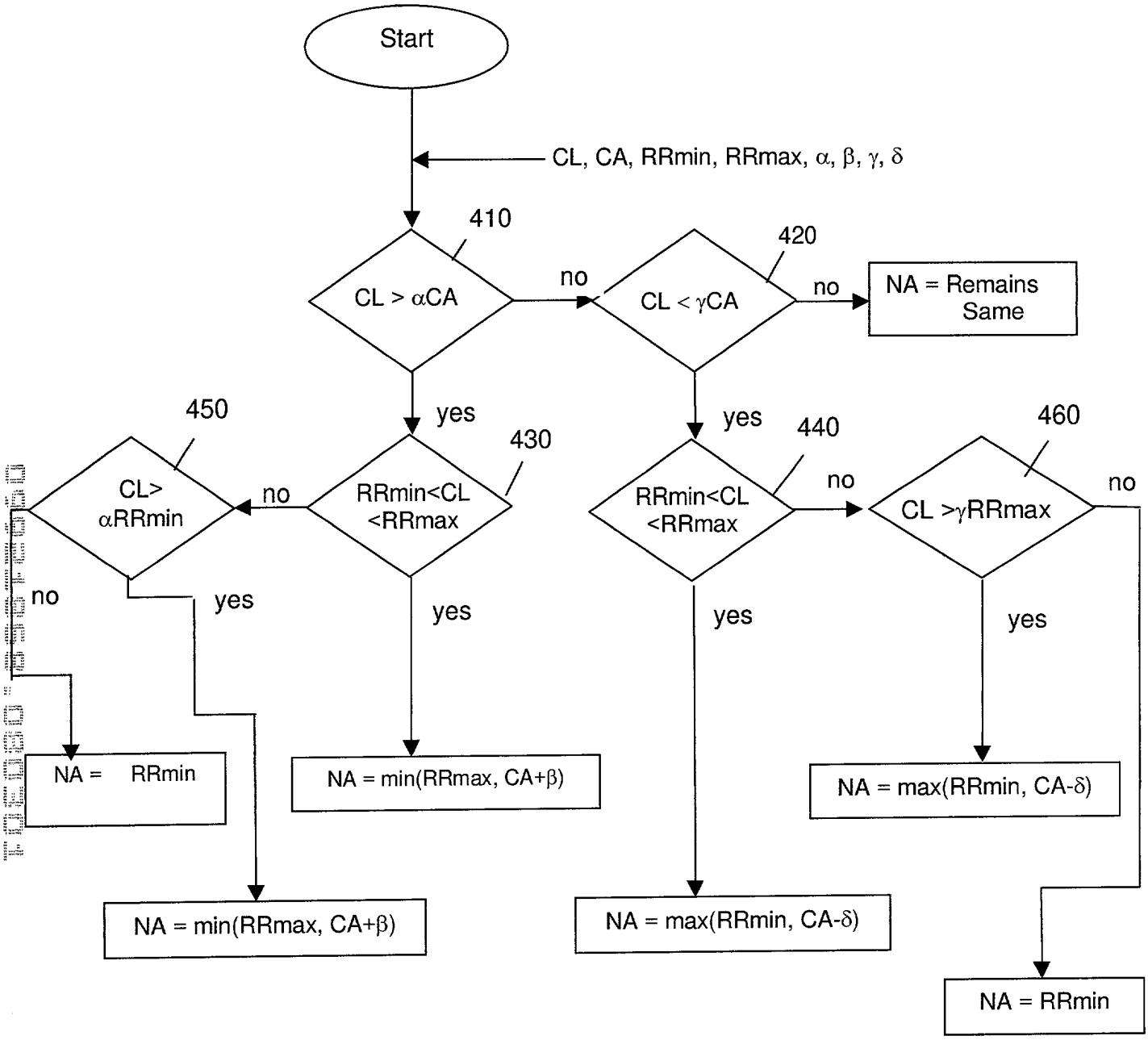

110
Load Distributor
130
Load Monitors
R
R
R
Load Factor per Resource Instance per user
Load Information
Hit rate
Aggregator
Level 1 Aggregator
322
120
Hit-Rate per Resource Class for each user
Hit-Rate and corresponding weight for each user and each Resource Class Instance
Level 2a Aggregator/local decision maker
324
Level 2b Aggregator
326
Change in Resource Requirements per Resource Class per user
Hit-Rate and correction factor per Resource Instance per user
Global Decision Maker
140

FIG 3



CL – Current Load
CA – Current Allocation
NA – New Allocation
RRmin – Minimum Resource Requirement
RRmax – Maximum Resource Requirement
α, β, γ, δ - configurable thresholds

FIG 4

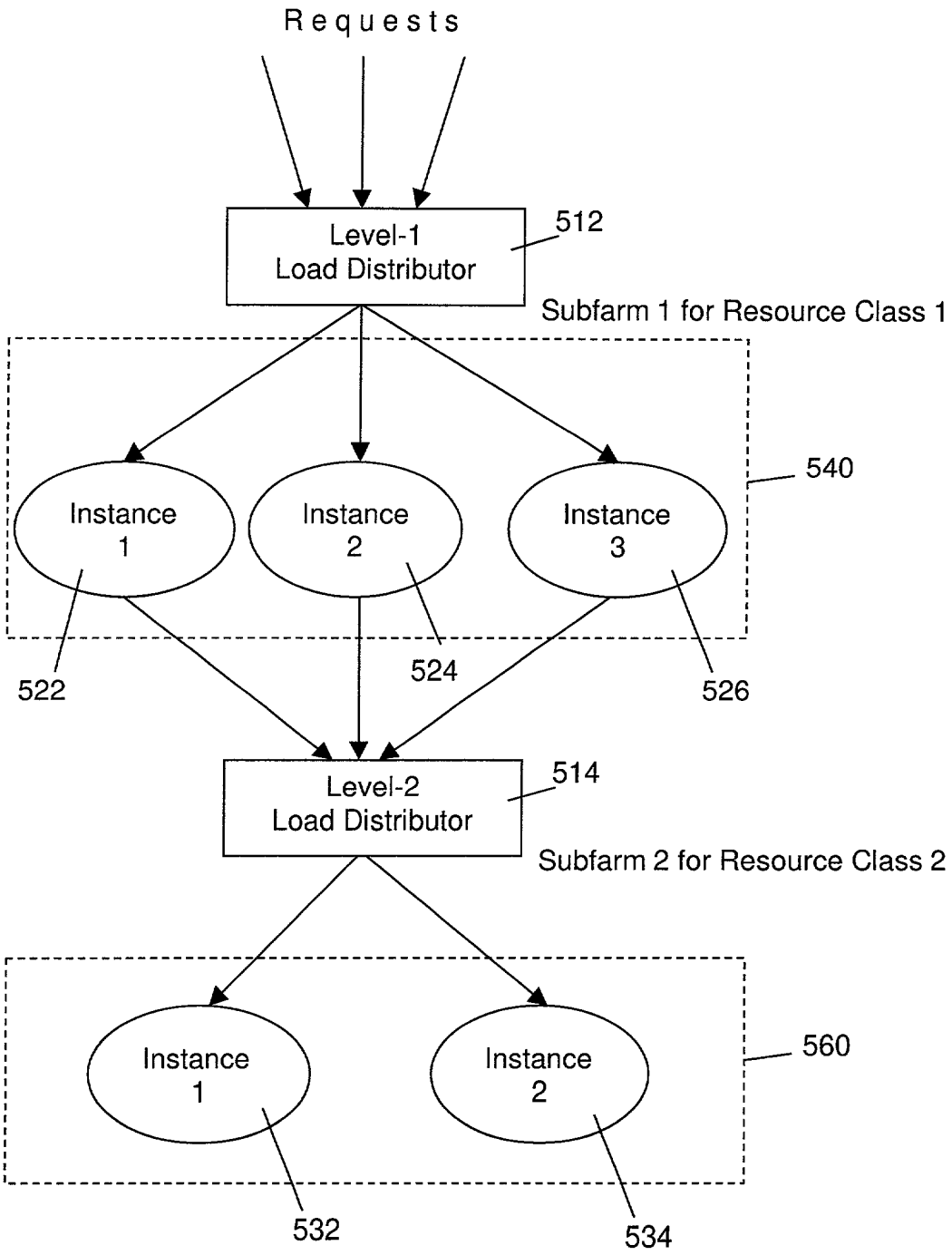

Requests
Level-1
Load Distributor
512
Subfarm 1 for Resource Class 1
Instance
1
Instance
2
Instance
3
540
522
524
526
Level-2
Load Distributor
514
Subfarm 2 for Resource Class 2
Instance
1
Instance
2
560
532
534

FIG 5

Legend: LD-n: n-th Load Distributor

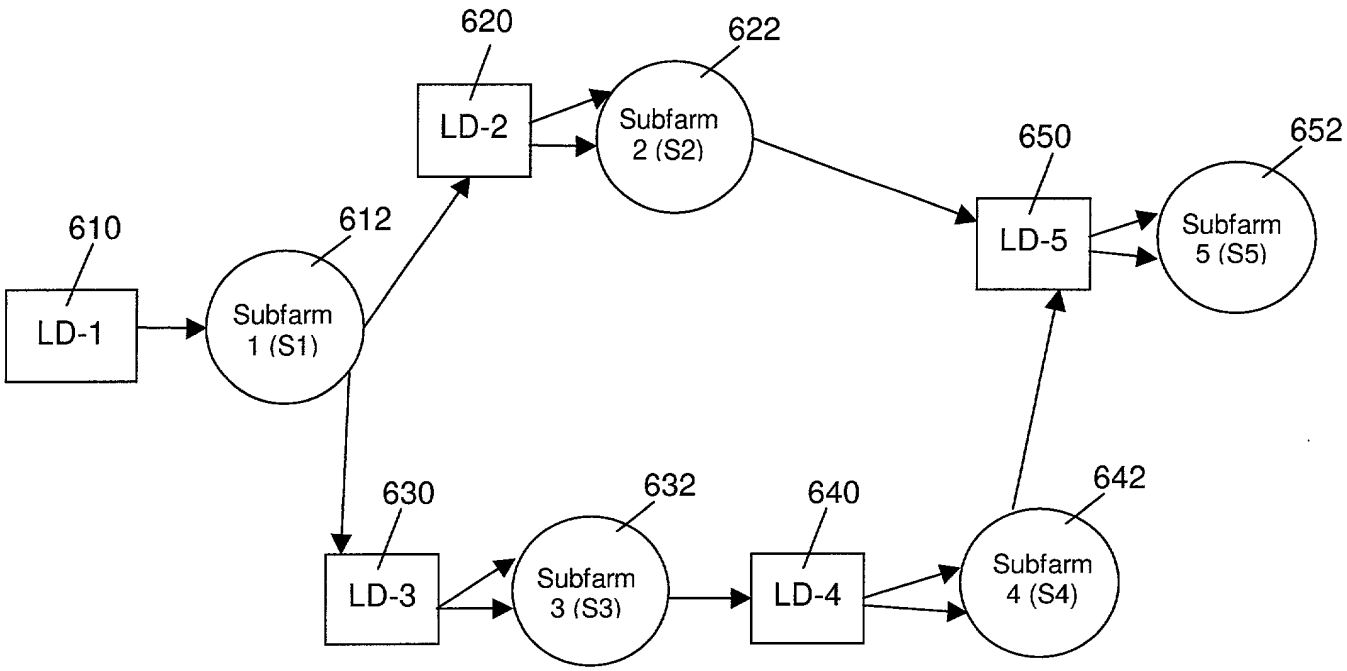


FIG 6

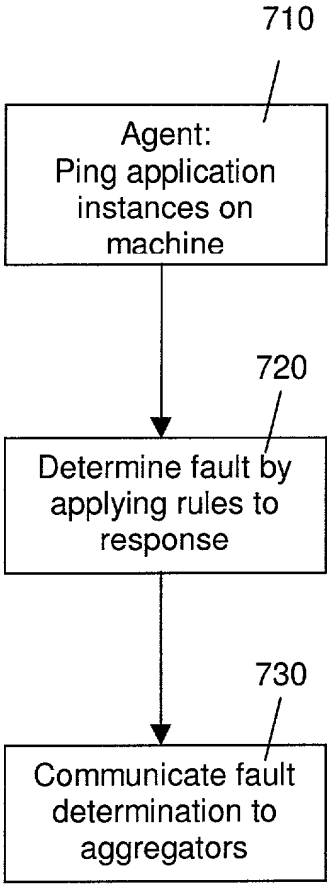

710
Agent:
Ping application instances on machine
720
Determine fault by applying rules to response
730
Communicate fault determination to aggregators

FIG 7A

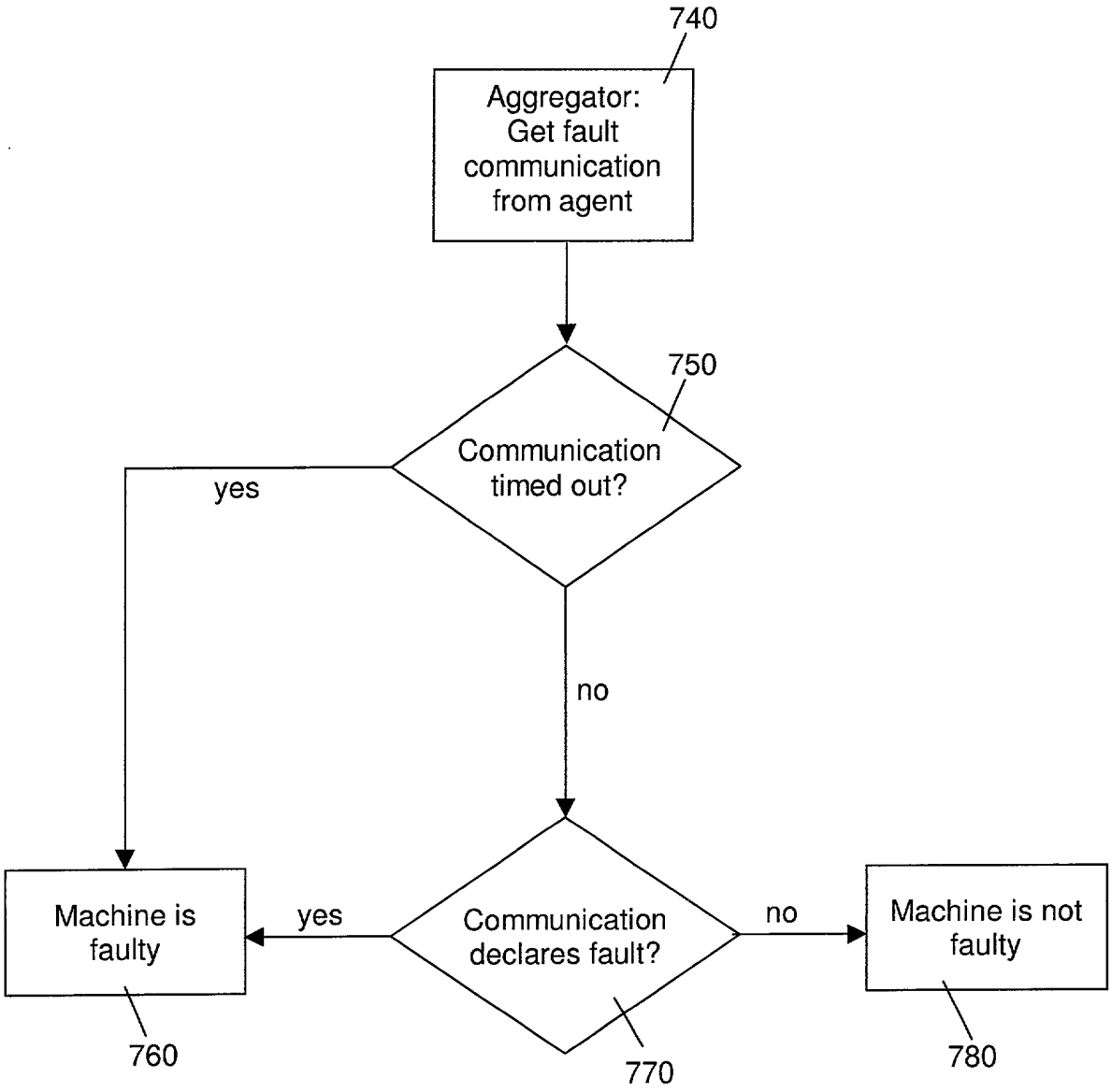
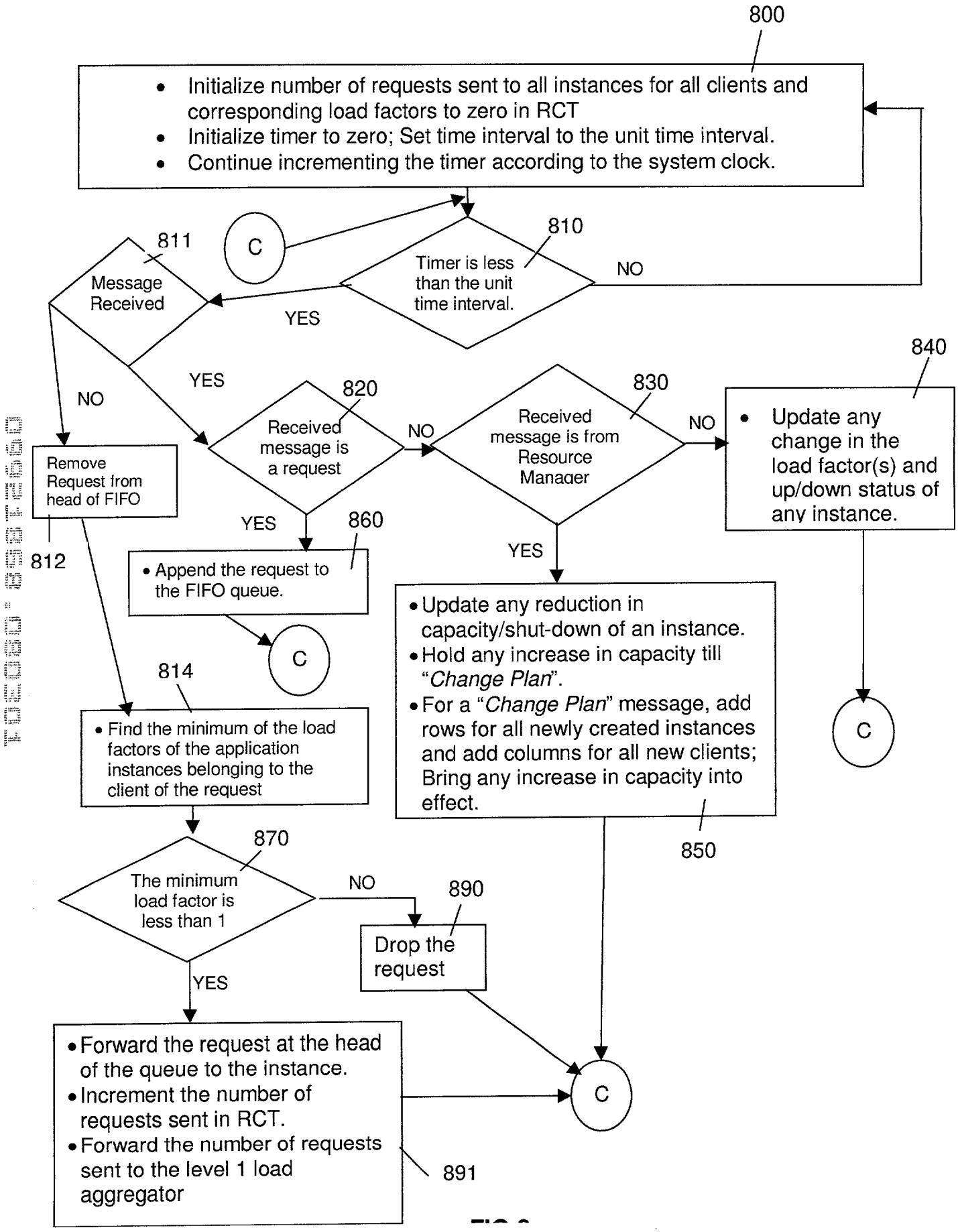

740
Aggregator: Get fault communication from agent
750
Communication timed out?
yes
no
Machine is faulty
yes
Communication declares fault?
no
Machine is not faulty
760
770
780

FIG 7B


800
• Initialize number of requests sent to all instances for all clients and corresponding load factors to zero in RCT
• Initialize timer to zero; Set time interval to the unit time interval.
• Continue incrementing the timer according to the system clock.
C
810
Timer is less than the unit time interval.
NO
YES
811
Message Received
NO
YES
820
Received message is a request
NO
830
Received message is from Resource Manager
NO
840
• Update any change in the load factor(s) and up/down status of any instance.
Remove Request from head of FIFO
812
YES
860
• Append the request to the FIFO queue.
C
YES
• Update any reduction in capacity/shut-down of an instance.
• Hold any increase in capacity till "Change Plan".
• For a "Change Plan" message, add rows for all newly created instances and add columns for all new clients; Bring any increase in capacity into effect.
850
C
814
• Find the minimum of the load factors of the application instances belonging to the client of the request
870
The minimum load factor is less than 1
NO
890
Drop the request
YES
• Forward the request at the head of the queue to the instance.
• Increment the number of requests sent in RCT.
• Forward the number of requests sent to the level 1 load aggregator
891
C

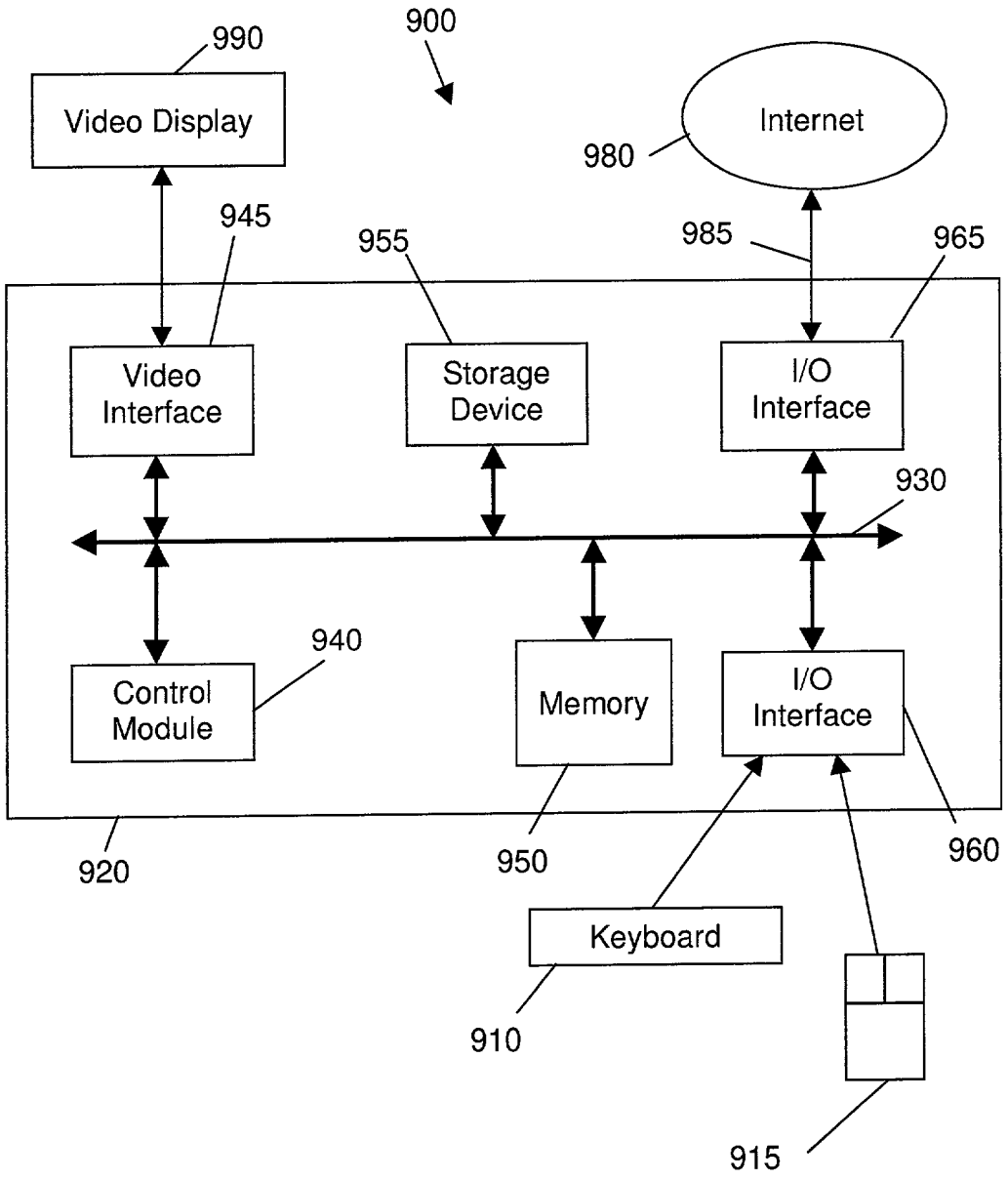

900
990
Video Display
945
955
980
Internet
985
965
Video Interface
Storage Device
I/O Interface
930
940
Control Module
Memory
I/O Interface
920
950
960
Keyboard
910
915

FIG 9